

## EL USO DE LA HERRAMIENTA “ANÁLISIS DE DATOS” DE EXCEL COMO COMPLEMENTO PARA LA ENSEÑANZA DE LA ESTADÍSTICA EN EL AULA

Álvaro Toledo San Martín  
alvaro.toledo@ubo.cl  
Universidad Bernardo O’Higgins, Chile

Tema: Uso de Tecnologías

Modalidad: Taller

Nivel educativo: Formación y actualización docente

Palabras clave: Excel – análisis de datos

### Resumen

*En este taller se mostrará la herramienta análisis de datos de Excel como complemento para el desarrollo de una clase de Estadística. En su primera parte se introducirá la implementación de ésta y funciones básica, para luego explorar las funciones asociadas a Estadística Descriptiva, Simulación de valores aleatorio e Inferencia. Se dará énfasis en la utilidad del software para realizar comprobaciones tales como el Teorema Central del Límite y el nivel de confianza en la construcción de un intervalo de confianza, entre otras.*

### Introducción

El uso de tecnologías de información y comunicación (TIC) es un complemento necesario para el desarrollo de una clase de Estadística, específicamente el uso de software se ha hecho fundamental para ejemplificar la teoría del análisis descriptivos de datos, la simulación de datos y su utilidad como complemento al concepto de variable aleatoria, la comprobación de teoremas relevante, la aplicación en inferencia, entre otros. En este sentido, son muchos los autores que consideran muy importante la actitud hacia el uso de software y el alcance que dan a estos los alumnos y docentes.

Por una parte, el profesorado ha enfrentado problemas para incorporar TIC en su práctica docente. Estos problemas se ven reflejados en una articulación inadecuada de estas tecnologías con las disciplinas que imparte, pobre contextualización, y en el mejor de los casos, un aprovechamiento insuficiente de su potencial. Por otra parte, el estudiantado es un factor aún más complejo para descifrar. En efecto, muchos estudiantes actuales no necesitan adaptarse a las nuevas tecnologías porque nacieron con ellas, lo que supone una

problemática aún mayor para autoridades educativas y profesores (Hernández, 2013, p. 166)

En esta misma línea, autores como Batanero (2009) destacan cómo la tecnología ha influido en la estadística y su enseñanza, siendo reconocida por la Internacional Association for Statistical Education (IASE) en los sucesivos Congresos Internacionales sobre la Enseñanza de la Estadística, donde se discute sobre el software disponible para la enseñanza, los cambios implicados en el contenido y la metodología, y el efecto en el aprendizaje y las actitudes de los alumnos. “Asimismo, se destaca cómo la tecnología ha reducido el tiempo de cálculo, permitiendo trabajar con aplicaciones reales en clase”. (Espinoza y Fernández, 2014, p. 90)

La estadística no es sólo una colección de conceptos y técnicas, sino sobre todo una forma de razonar. Es necesaria en la mayoría de los ámbitos, en cualquier ciencia o cualquier trabajo (Martín, Cabero y De Paz, 2008). Bajo esta premisa se plantea entonces el desarrollo de este taller y el software seleccionado es Microsoft Excel, el porque de su elección es simple, está disponible en la mayoría de los laboratorios de computación en colegios o Universidades. “Microsoft Excel es la hoja de cálculo más popular que se utiliza para almacenar información en columnas y filas, que luego pueden ser organizadas y/o procesadas” (Giles, 2002, p.2). Muchos autores consideran Microsoft Excel como una excelente herramienta para la enseñanza de la estadística. “Una ventaja importante de la hoja de cálculo Excel es que se ha convertido en un software estándar en los entornos de enseñanza, profesionales y familiares . Excel tiene una interfaz amigable y es fácil usar” (Cao y Naya, 2013, p.2).

Para este taller se utilizarán funciones de Excel relacionadas con la confección de tablas y gráficos dinámicos, algunas funciones lógicas para el tratamiento previo de variables de estudio y el complemento “Análisis de datos” como herramienta para profundizar en elementos de estadística descriptiva, probabilidad e inferencia.

## Desarrollo del taller

Tal como se mencionó anteriormente, Excel cuenta con un complemento denominado “Análisis de datos” tal herramienta permite trabajar datos de forma descriptiva, simular datos e incluso realizar inferencia estadística en base a intervalos de confianza y test de hipótesis entre otras opciones. Como primer paso, será entonces mostrar a los participantes como activar tal complemento desde la planilla de Excel para luego mostrar a grandes rasgos el menú de opciones disponibles en la ventana de Análisis de datos y comenzar a trabajar descriptivamente datos. La primera actividad se enfocará en las funciones “Estadística Descriptiva” e “Histograma” previo trabajo con tablas dinámicas y alguna funciones útiles para transformar variables (categorizar o creación de variables nuevas) luego, se realizarán algunos métodos descriptivos bivariados: tablas de doble entrada, cálculo de covarianza y correlación y análisis de regresión lineal simple para esto se utilizarán las funciones “Covarianza”, “Coeficiente de correlación” y “Regresión”. Como segunda parte del taller se utilizará la función “Generación de números aleatorios” la cual se revisará en profundidad por su utilidad para mostrar el resultado del Teorema Central del Límite, Simulación de Montecarlo e intervalos de confianza. Como parte final se realizarán actividades enfocadas en mostrar el resultado del Teorema Central del Límite (distribución Normal de la media de un conjunto de variables aleatoria independientes e idénticamente distribuidas), uso de la simulación de Montecarlo para generar valores aleatorios de una distribución exponencial (no incluida en Excel) y la verificación del concepto de confianza en la construcción de un Intervalo de confianza. Se explorarán además las opciones relacionas con test (prueba) de hipótesis para medias de dos poblaciones y de razón de varianzas.

## Algunas actividades propuestas

1. Análisis de datos. Problema extraído del texto Estadística para Administración y Economía (Anderson, D., Sweeney, D. y Williams, T., 2008, pp.66):

Las tiendas Pelican, una división de National Clothing, es una cadena de tiendas

de ropa para mujer que tiene sucursales por todo Estados Unidos. Hace poco la tienda realizó una promoción en la que envió cupones de descuento a todos los clientes de otras tiendas de National Clothing. Los datos obtenidos en una muestra de 100 pagos con tarjeta de crédito en las tiendas Pelican durante un día de la promoción se presentan en el archivo titulado PelicanStores

El modo de pago Proprietary card se refiere a pagos realizados usando una tarjeta de crédito de National Clothing. A los clientes que hicieron compras usando un cupón de descuento se les denomina aquí promocionales y a quienes hicieron sus compras sin emplear cupón de descuento se les denomina regulares. Como a los clientes de las tiendas Pelican no se les enviaron cupones promocionales, los directivos consideran que las ventas hechas a quienes presentaron un cupón de descuento son ventas que de otro modo no se hubieran hecho. Es claro que Pelican espera que los clientes promocionales continúen comprando con ellos.

La mayor parte de las variables se explican por sí mismas, pero dos de las variables deben ser aclaradas.

Artículos: El número total de artículos comprados

Ventas netas: Cantidad total cargada a la tarjeta de crédito

Los directivos de Pelican desean emplear estos datos muestrales para tener información acerca de sus clientes y para evaluar la promoción utilizando los cupones de descuento.

Informe para los directivos

Emplee los métodos tabulares y gráficos de la estadística descriptiva para ayudar a los directivos de Pelican a elaborar un perfil de sus clientes y a evaluar la promoción. Su informe debe contener, por lo menos, lo siguiente:

- i) Distribuciones de frecuencia porcentual de las variables clave.
- ii) Una gráfica de barras o una gráfica de pastel que muestre el número de



clientes correspondiente a cada modo de pago.

- iii) Una tabulación cruzada con el tipo de cliente (regular o promocional) frente a ventas netas. Haga un comentario sobre las semejanzas o diferencias que observe.
- iv) Un diagrama de dispersión para investigar la relación entre ventas netas y edad del cliente.

## 2. Aplicación del método de Montecarlo

Utilice los resultados del método de Montecarlo discutidos en el taller para simular 1000 datos provenientes para una variable aleatoria  $X$  con función de densidad dada por:

$$f(x) = 3x^2, 0 < x < 1$$

Luego, determine la proporción de valores inferiores a 0,7. Compare este valor con su valor teórico.

## 3. Visualización del Teorema Central del Límite.

Considere el experimento aleatorio de extraer de una urna 3 esferas con reposición. La urna está compuesta de 5 esferas azules y 3 esferas rojas. Nos interesa el número de esferas azules extraídas.

- i) Simule el experimento descrito en el enunciado.
- ii) Repita el experimento aleatorio 1000 veces
- iii) Realice una tabla de frecuencias para el punto anterior, luego grafique. Comente.
- iv) Simule 500 veces la repetición de 1000 veces del experimento aleatorio.

- v) Determine el promedio para cada una de las 500 columnas obtenidas anteriormente.
  - vi) Grafique los promedios. ¿Qué observa?
4. El papel de la confianza en el intervalo de confianza.

Considere una variable aleatoria  $X$ : Estatura de un latinoamericano (metros).

- i) Simule 1000 datos de una distribución Normal con media 1,68 y desviación estándar 0,05 (¿qué implica esta desviación estándar?). Repita esta simulación 200 veces.
- ii) Para cada una de las repeticiones determine su respectivo intervalo de confianza para la media poblacional (puede considerar el caso de varianza poblacional conocida).
- iii) Finalmente, determine la proporción de intervalos de confianza que contienen a la media poblacional del enunciado.

### Referencias bibliográficas

- Anderson, D., Sweeney, D. y Williams, T. (2008). *Estadística para administración y Economía*. México: Cengage Learning.
- Batanero, C. (2009). Retos para la formación estadística de los profesores. II Encontro de Probabilidade e Estatística na Escola. Universidade do Minho, 2009, Braga, Portugal.  
<http://www.ugr.es/~batanero/pages/ARTICULOS/Formprofesores.pdf>.  
Consultado el 30/01/2016.
- Cao, R. y Naya, S. (2010). The use of statistical software to teach nonparametric curve estimation: from Excel to R. 8th International Conference on Teaching Statistics. Ljubljana, Slovenia. Disponible en  
[http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8\\_4B1\\_CAO.pdf](http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_4B1_CAO.pdf).  
Consultado el 23/01/2016.

- Espinoza C., Fernández J. (2014). Importance of statistical software teaching and learning at the university of Carabobo (Venezuela). *Aula de encuentro* 16(1), 89-102.
- Giles, O. (2002). Using excel to teach statistics in New Zealand secondary schools. ICOTS 2002. [http://iase-web.org/documents/papers/icots6/7g1\\_gile.pdf](http://iase-web.org/documents/papers/icots6/7g1_gile.pdf). Consultado el 20/01/2016.
- Hernández, S., Cuevas, J. (2013). Programas informáticos de uso libre y su aplicación en la enseñanza de la Estadística. *Revista Investigación Operacional* 34(2), 166-174.
- Martín Martín, Q., Cabero, M. T. & de Paz Santana, Y. (2008). *Tratamiento estadístico de datos con SPSS*. España: Thomson.